Function Approximation through Fuzzy Systems Using Taylor Series Expansion-Based Rules: Interpretability and Parameter Tuning

Luis Javier Herrera[1], Héctor Pomares[1], Ignacio Rojas[1], Jesús González[1], and Olga Valenzuela[2]

[1] University of Granada, Department of Computer Architecture and Technology, E.T.S. Computer Engineering, 18071 Granada, Spain
http://atc.ugr.es

[2] University of Granada, Department of Applied Mathematics, Science faculty, Granada, Spain

Abstract. In this paper we present a new approach for the problem of approximating a function from a training set of I/O points using fuzzy logic and fuzzy systems. Such approach, as we will see, will provide us a number of advantages comparing to other more-limited systems. Among these advantages, we may highlight the considerable reduction in the number of rules needed to model the underlined function of this set of data and, from other point of view, the possibility of bringing interpretation to the rules of the system obtained, using the Taylor Series concept. This work is reinforced by an algorithm able to obtain the pseudo-optimal polynomial consequents of the rules. Finally the performance of our approach and that of the associated algorithm are shown through a significant example.

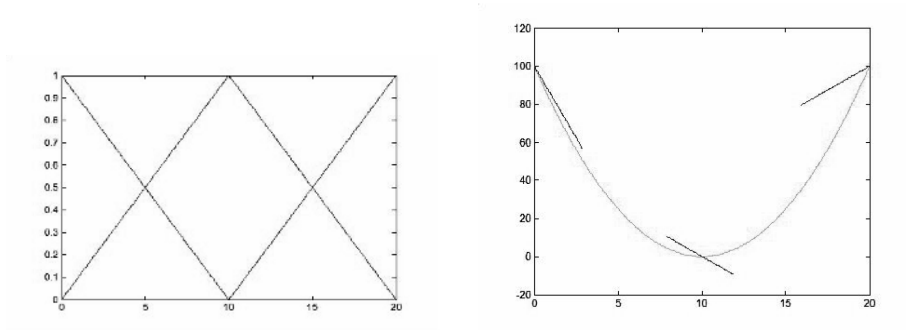
1 Introduction

The Function Approximation problem deals with the estimation of an unknown model from a data set of continuous input/output points; the objective is to obtain a model from which to get the expected output given any new input data.

Fuzzy Logic on the other hand is one of the three roots of soft-computing; it has been successfully applied to several areas in scientific and engineering sectors, due to its broad number of benefits. The simplicity of the model and its understandability, while encapsulating complex relations among variables is one of the keys of the paradigm. The other main characteristic is its capability to interpret the model, for example through the use of linguistic values to bring meaning to the variables involved in the problem.

Many authors have dealt with Fuzzy logic and Fuzzy Systems for function approximation from an input/output data set, using clustering techniques as well as grid techniques, obtaining in general good enough results. Specifically, the TSK model [7] fits better to these kind of problems due to it's computational capability.

R. Monroy et al. (Eds.): MICAI 2004, LNAI 2972, pp. 508–516, 2004.



(a) (b)

Fig. 1. a) MF distribution used for this example. Target function: y = (x - 10)2 **b)** Original function + model output + linear submodels for each of the three rules using a TSK model of order 1. We see how the global output of the TSK fuzzy system is eye-indistinguishable from the actual output, but no interpretation can be given to the three linear sub-models.

The fuzzy inference method proposed by Takagi, Sugeno and Kang, which is known as the TSK model in fuzzy systems field, has been one of the major issues in both theoretical and practical research for fuzzy modelling and control. The basic idea is the subdivision of the input space into fuzzy regions and to approximate the system in each subdivision by a simple model.

The main advantage of the TSK model is its representative power, capable of describing a highly complex nonlinear system using a small number of simple rules. In spite of this, the TSK systems suffer from the lack of interpretability, which should be one of the main advantages of fuzzy systems in general. While the general performance of the whole TSK fuzzy system give the idea of what it does, the sub-models given by each rule in the TSK fuzzy system can give no interpretable information by themselves [1]. See *Fig* 1.

Therefore, this lack of interpretability might force any researcher not to use the TSK models in problems where the interpretability of the obtained model and corresponding sub-models is a key concept.

Apart from the interpretability issue the number of rules for a working model is also a key concept. For control problems, grid-based fuzzy systems are preferable since they cover the whole input space (all the possible operation regions in which the plant to control can be stated during its operation). Nevertheless, for Mamdani fuzzy systems or even for TSK fuzzy systems of order 0 or 1, although getting pseudo-optimal solutions, they usually need an excessive number of rules for a moderated number of input variables.

In this paper we propose the use of high order TSK rules in a grid based fuzzy system, reducing the number of rules, while keeping the advantages of the grid-based approach for control and function approximation. Also to keep the interpretability of the model obtained, we present a small modification for the consequents of the high order TSK rules in order to provide the interpretability for each of the sub-models (rules) that compose the global system.

The rest of the paper is organized as follows: Section 2 presents high order TSK rules with an algorithm to obtain the optimal coefficients of the rule consequents. Section 3 provides an introduction to the Taylor Series Expansion, concept that will provide the key for the interpretability issue, commented in Section 4. Finally, in section 5 it is provided a whole example that demonstrates the suitability and goodness of our approach.

2 High-Order TSK Fuzzy Rules

The fuzzy inference system proposed by Takagi, Sugeno and Kang, known as the TSK model in the fuzzy system literature, provides a powerful tool for modelling complex nonlinear systems. Typically, a TSK model consists of IF-THEN rules that have the form:

$$\begin{aligned} R^k : \text{IF } x_1 \text{ is } A_1^k \text{ AND } \ldots \text{ AND } x_n \text{ is } A_n^k \text{ THEN} \\ y = \alpha_0^k + \alpha_1^k x_1 + \ldots + \alpha_n^k x_n \end{aligned} \tag{1}$$

where the A_i^k are fuzzy sets characterized by membership functions $A_i^k(x)$, α_j^k are real-valued parameters and x_i are the input variables.

A Sugeno approximator comprises a set of TSK fuzzy rules that maps any input data $\vec{x} = [x_1, x_2, \ldots, x_n]$ into its desired output $y \in \mathbb{R}$. The output of the Sugeno approximator for any input vector x, is calculated as follows:

$$F(x) = \frac{\sum_{k=1}^{K} \mu_k(x) y_k}{\sum_{k=1}^{K} \mu_k(x)} \tag{2}$$

Provided that $\mu_k(x)$is the activation value for the antecedent of the rule k, and can be expressed as:

$$\mu_k(x) = A_1^k(x_1) A_2^k(x_2) \ldots A_n^k(x_n) \tag{3}$$

The main advantage of the TSK model is its representative power; it is capable of describing a highly nonlinear system using a small number of rules. Moreover, since the output of the model has an explicit functional expression form, it is conventional to identify its parameters using some learning algorithms. These characteristics make the TSK model very suitable for the problem of function approximation; a high number of authors have successfully applied TSK systems for function approximation. For example, many well-known neuro-fuzzy systems such as ANFIS [4] have been constructed on the basis of the TSK model.

Nevertheless very few authors have dealt with high-order TSK fuzzy systems. Buckley [5] generalized the original Sugeno inference engine by changing the form of the consequent to a general polynomial, that is:

$$R^k : \text{IF } x_1 \text{ is } A_1^k \text{ AND } \ldots \text{ AND } x_n \text{ is } A_n^k \text{ THEN } y = Y_k(x) \tag{4}$$

where $Y_k(x)$ is a polynomial of any order. Taking order 2, it can be expressed as:

$$Y_k(\vec{x}) = w_0^k \cdot \vec{x} + \frac{1}{2}\vec{x}^T W^k \vec{x} \tag{5}$$

Where w_0 is a scalar, $\vec{w}$ is a column vector of coefficients with dimension n (one per each input variable) and W is a triangular matrix of dimensions $n{\times}n$, (W_{ij} = coefficient for quadratic factor $x_i * x_j$, $i = 1 \ldots n$, $j = i \ldots n$).

Now that we have defined how a TSK fuzzy system can be adapted to work with high-order rules, let's see, given a set of input/output data, and a configuration of membership functions for the input variables, how to adapt the consequents of the rules so that the TSK model output optimally fits the data set D. The Least Square Error (LSE) algorithm will be used for that purpose. LSE tries to minimize the error function:

$$J = \sum_{m \in D} (y_m - F(x))^2 \tag{6}$$

where F is the output of the TSK fuzzy system as in (2). Setting to 0 the first derivative (7, 8) of each single parameter (w_0 and each component of $\vec{w}$ and W) will give us a system of linear equations from which to obtain the optimal values of the parameters.

$$\frac{\partial J}{\partial w_{si}} = 2 \sum_{m \in D} \left(y_m - \frac{\sum_{k=1}^{K} \mu_k(x_m)\left(w_0^k + \vec{w}^k \vec{x}_m + \frac{1}{2}\vec{x}^T W^k \vec{x}_m\right)}{\sum_{k=1}^{K} \mu_k(x_m)} \right) \cdot \frac{\mu_s(x_m) f_{w_{si}}(x_m)}{\sum_{k=1}^{K} \mu_k(x_m)} \tag{7}$$

$$\begin{aligned} &\sum_{m \in D} \frac{y_m \cdot \mu_s(x_m) \cdot f_{w_{si}}(x_m)}{\sum_{k=1}^{K} \mu_k(x_m)} = \\ &= \sum_{k=1}^{K} \sum_{i} w_{ki} \sum_{m \in D} \frac{\mu_k(x_m) \cdot f_{w_{ki}} \cdot \mu_s(x_m) \cdot f_{w_{si}}(x_m)}{\sum_{j=1}^{K} \mu_j(x_m)} \end{aligned} \tag{8}$$

Where w_{ki} -rule k, coefficient i- is the coefficient we are differentiating in each case (w_0 or any component of W or $\vec{w}$), and f_{wi} is the partial derivative of the consequent of rule k with respect to w_i, i.e., 1 for the 0-order coefficient w_0, x_i for every first-order coefficient w_i, or $x_p \cdot x_j$ for every second-order coefficient w_{pj} of W.

Once we have the system of linear equations, it only remains to obtain the optimal solution for all the coefficients of every rule. The Orthogonal Least-Square (OLS) method [6] will guarantee a single optimal solution obtaining the values for the significant coefficients while discarding the rest. We reject therefore the problems due to the presence of redundancy in the activation matrix.

Once that we have already reviewed the type of rules that we are going to operate with, now let's review the "lack of interpretability curse" that suffer TSK

Fuzzy Systems as we saw in Section 1. As polynomials are not easy interpretable as consequents of the rules, we will give now the key for the interpretability for our Taylor-Series based rules.

3 Taylor Series-Based Fuzzy Rules (TSFR)

Let $f(x)$ be a function defined in an interval with an intermediate point a, for which we know the derivatives of all orders. The first order polynomial:

$$p_1(x) = f(a) + f'(a)(x-a) \tag{9}$$

has the same value as $f(x)$ in the point $x = a$ and also the same first order derivative at this point. Its graphic representation is a tangent line to the graph of $f(x)$ at the point $x = a$.

Taking also the second derivative for $f(x)$ in $x = a$, we can build the second order polynomial

$$p_2(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 \tag{10}$$

which has the same value as $f(x)$ at the point $x = a$, and also has the same values for the first and second derivative. The graph for this polynomial in $x = a$, will be more similar to that of $f(x)$ in the points in the vicinity of $x = a$. We can expect therefore that if we build a polynomial of nth order with the n first derivatives of $f(x)$ in $x = a$, that polynomial will get very close to $f(x)$ in the neighbourhood of $x = a$.

Taylor theorem states that if a function $f(x)$ defined in an interval has derivatives of all orders, it can be approximated near a point $x = a$, as its Taylor Series Expansion around that point:

$$\begin{aligned} f(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \dots \\ + \frac{1}{n!}f^{(n)}(a)(x-a)^n + \frac{1}{(n+a)!}f^{(n+1)}(c)(x-a)^{n+1} \end{aligned} \tag{11}$$

where in each case, c is a point between x and a.

For n-dimensional purposes, the formula is adapted in the following form:

$$\begin{aligned} f(\vec{x}) = f(\vec{a}) + (\vec{x}-\vec{a})^T \left[\frac{\partial f}{\partial \vec{x}_i}(\vec{a})\right]_{i=1\dots n} + \frac{1}{2}(\vec{x}-\vec{a})^T W(\vec{x}-\vec{a}) + \\ \frac{1}{3!}W^3(\vec{x}-\vec{a}, \vec{x}-\vec{a}, \vec{x}-\vec{a}) + \dots \end{aligned} \tag{12}$$

where W is a triangular matrix of dimensions $n \times n$, and W^s is a triangular multi-linear form in s vector arguments $v^1, \dots, v^s$.

Taylor series open a door for the approximation of any function through polynomials, that is, through the addition of a number of simple functions. It

is therefore a fundamental key in the field of Function Approximation Theory and Mathematical Analysis. Taylor Series Expansion will also provide us a way to bring interpretation to TSK fuzzy systems by taking a certain type of rules consequents and antecedents, as we will now see.

As noted in [3] we will use input variables in the antecedents with membership functions that form an Orderly Local Membership Function Basis (OLMF). The requirements that a set of membership functions for a variable must fulfil to be an OLMF basically are:

- Every membership function extreme point must coincide with the centre of the adjacent membership function.
- The *n-th* derivative of the membership function is continuous in its whole interval of definition.
- The *n-th* derivative of the membership function vanishes at the centre and at the boundaries.

The main advantage of using this kind of membership functions is the differentiability of the output of the TSK fuzzy system. This is not possible when we have triangular or trapezoidal membership functions, since the derivative at the centres of the membership functions does not exist, therefore not having a differentiable fuzzy system output.

These OLMF bases also have the *addition to unity property*: the addition of the activations of all the rules is always equal to unity for any point inside the input domain in a TSK fuzzy system that keeps the OLMF basis restrictions. Therefore the output of the TSK fuzzy system can be expressed as:

$$F(x) = \sum_{k=1}^{K} \mu_k(x) y_k \tag{13}$$

Then the OLS method cited in Section 2 will work well for the given system, and can identify the optimal coefficients without needing another execution of the algorithm as noticed in [6].

Finally, given that the input variables have a distribution of membership functions that form a OLMF basis, we will use high-order TSK rules in the form (4), but where the polynomial consequents are in the form:

$$Y_k(\vec{x}) = w_0^k \cdot \vec{w}^k(\vec{x} - \vec{a}_k) + \frac{1}{2}(\vec{x} - \vec{a}_k)^T \cdot W^k \cdot (\vec{x} - \vec{a}_k) \tag{14}$$

being $\vec{a}_k$ the centre of rule k, therefore forming a Taylor Series Expansions around the centres of the rules.

4 Interpretability Issues

It can be demonstrated [3] that given a Sugeno approximator $F(x)$ such that:

- 1) the input variables membership functions form a set of OLMF basis of order m (being the m-th derivative continuous everywhere);

– 2) the consequent-side is written in the rule-centred form shown in (4) and (14) and the polynomials $Y_k(x)$ are of degree n.

Then for $n \leq m$, every $Y_k(x)$ can be interpreted as a truncated Taylor series expansion of order n of $F(x)$ about the point $\vec{x} = \vec{a}_k$, the centre of the kth rule.

Supposing therefore that we have a method to obtain the optimal Taylor-Series Based TSK rules consequents coefficients for function approximation, given a data set and a membership function distribution that form a set of OLMF basis, we can interpret then the consequents of the rules $Y_k(x)$ as the truncated Taylor series expansion around the centres of the rules of the output of the system. This system also provides a pseudo-optimal approximation to the objective function. In the limit case where the function is perfectly approximated by our system, the rule consequents will coincide with the Taylor Series expansions of that function about centre of each rule, having reached total interpretability and total approximation.

In [3], Marwan Bikdash used directly the (available) Taylor Series Expansion of the function around the rule centres, for each rule, to approximate the function with the TSK fuzzy system. Notice that these rule consequents, though having strong interpretability, are not the optimal consequents in the least squares sense. Please note that the Taylor Series Expansion is an approximation for a function in the vicinity of the reference point. Therefore even using a high number of MFs, the error obtained by the method in [3] is seldom small enough (compared to a system with similar complexity with consequents optimized using LSE) and therefore the system output barely represent a good approximation of the data we are modelling.

In this paper we also suppose that the only information we have from the function to approximate are the input/output points in the initial dataset. No information is given of the derivatives of the function w.r.t. any point. Also, there is no accurate way to obtain the derivatives from the training points to perform the approximation as the method in [3] required.

5 Simulations

Consider a set of 100 randomly chosen I/O data from the 1-D function [2]:

$$F(x) = e^{-5x} sin(2\pi x) \ \in \ [0, 1] \tag{15}$$

Let's try now to model those data using a fuzzy system with 5 membership functions for the single input variable x forming a OLMF basis and rule consequents of the form given by (14), being Y_k a order-2 polynomial.

The five rules obtained after the execution of the LSE algorithm using OLS are the following:

$$\begin{array}{l} \text{IF } x \text{ is } A_1 \text{THEN } y = -26.0860x^2 + 5.3247x + 0.0116 \\ \text{IF } x \text{ is } A_2 \text{THEN } y = -1.6235(x - 0.25) + 0.2882 \\ \text{IF } x \text{ is } A_3 \text{THEN } y = 4.3006(x - 0.5)^2 - 0.5193(x - 0.5) \\ \text{IF } x \text{ is } A_4 \text{THEN } y = -1.1066(x - 0.75)^2 + 0.1780(x - 0.75) - 0.0238 \\ \text{IF } x \text{ is } A_5 \text{THEN } y = 0 \end{array} \tag{16}$$

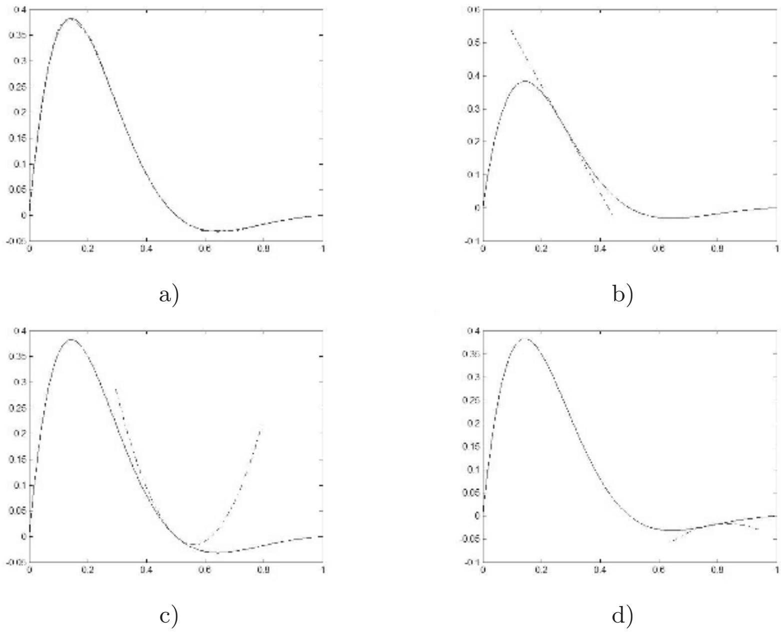


Fig. 2. a) Original function (solid line) and Taylor Series Expansion based Fuzzy System (dotted line). We see that for only 5 membership functions, the output of the system is very similar to the original function. NRMSE = 0.0154. b) Original function + model output + second membership function consequent (centered at x=0.25). c) Original function + model output + third membership function consequent (centered at x=0.5). d) Original function + model output + fourth membership function consequent (centered at x=0.75). We see clearly how these polynomials come closer to the Taylor Series Expansion around the centre of the rules of the fuzzy system output.

The interpretability comes from the fact that the function in the points near to each center of the five rules is extremely similar to the polynomial output of the rules as shown in Figure 2. These polynomials are kept expressed as Taylor Series Expansions of the function in the points in the vicinity of the centres of the rules. The system is therefore fully interpretable and also brings some more advantages as noticed below.

Figure 2 also shows clearly that the LSE finds the optimal consequents coefficients for the given input/output data set. Also it must be noted that for only five rules (one per each membership function), the error obtained is sensibly low. If we compare the system obtained for the same number of rules with a TSK fuzzy system with constant consequents, we see that the error obtained (NRMSE = 0.3493) is very high comparing to our Taylor Series Expansion based fuzzy system (NRMSE = 0.0154).

It should be remembered that the Normalized Root-Mean Square Error (NRMSE) is defined as:

$$NRMSE = \sqrt{\frac{\overline{e^2}}{\sigma_y^2}} \tag{17}$$

where σ_y^2 is the variance of the output data, and $\overline{e^2}$ is the mean-square error between the system and the dataset D output.

Also comparing using the same number of parameters, that is, 5 rules for our system, 15 rules for constant consequents TSK rules, we observe that the error obtained by our Taylor-Based rules system is much lower (NRMSE = 0.0154) than for constant consequent rule system (NRMSE = 0.0635).

6 Conclusions

In this paper we have presented a very interesting approach to the problem of function approximation from a set of I/O points utilizing a special type of fuzzy systems. Using an Orderly Local Membership Function Basis (OLMF) and Taylor Series-Based Fuzzy Rules, the proposed fuzzy system has the property that the Taylor Series Expansion of the defuzzified function around each rule centre coincides with that rule's consequent. This endows the proposed system with both the approximating capabilities of TSK fuzzy rules through the use of the OLS algorithm, and the interpretability advantages of pure fuzzy systems.

Acknowledgements. This work has been partially supported by the Spanish CICYT Project DPI2001-3219.

References

1. Yen, J., Wang, L., Gillespie, C.W.: Improving the Interpretability of TSK Fuzzy Models by Combining Global Learning and Local Learning. IEEE Trans. Fuz. Syst., Vol.6, No.4.(1998) 530–537
2. Pomares, H., Rojas, I., Gonzalez, J., Prieto, A.: Structure Identification in Complete Rule-Based Fuzzy Systems. IEEE Trans. Fuzz. Vol.10, no. 3. June (2002) 349–359
3. Bikdash, M.: A Highly Interpretable Form of Sugeno Inference Systems. IEEE Trans. Fuz. Syst., Vol.7, No.6. (1999) 686–696
4. Jang, J.-S R.: ANFIS: Adaptative-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern., vol. 23. (1993) 665–685
5. Buckley, J.J.: Sugeno-type Controllers are Universal Controllers. Fuzzy Sets. Syst., Vol.25, (1993) 299–303
6. Yen, J., Wang, L.: Simplifying Fuzzy Rule-Based Models Using Orthogonal Transformation Methods. IEEE Trans. Syst. Man Cybern., part B: cybernetics, Vol. 29, No.1. (1999) 13–24
7. Takagi, T.,Sugeno, M.: Fuzzy identification of systems and its applications to modelling and control. IEEE Trans. Syst. Man and Cyber., vol.15.(1985) 116–132